

GC-MS solution 工作站数据平滑的几种方法介绍

谯应召

(山东化工研究院, 山东 济南 250013)

摘要: GC-MS solution 色谱质谱数据工作站实现数据平滑的集中方法与基本原理解释。

关键词: 数据平滑

1 案例介绍

参见 <https://bbs.instrument.com.cn/topic/7642927>。

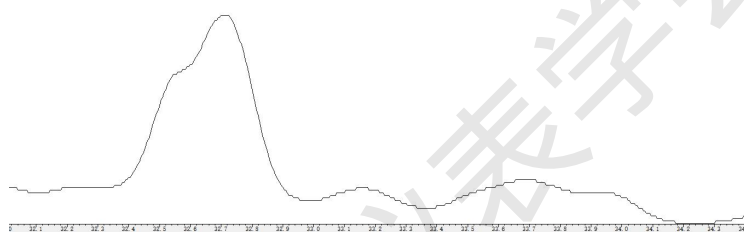


图1 MRM 方式下的 TIC 色谱图

并且用户自述在 Scan 方式下的数据，异构体出峰比较正常，如图 2 所示。

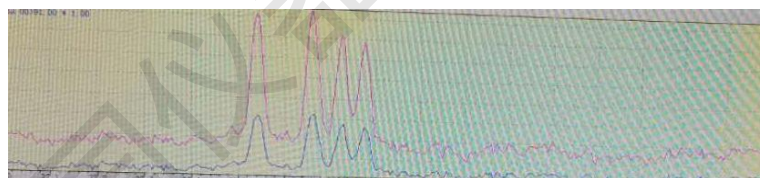


图2 Scan 方式下的 TIC 色谱图

2 故障处理

根据故障现象，怀疑问题出在数据平滑参数。

打开用户数据文件，检查色谱图定性积分参数，发现数据的“平滑方法”选择了较大的平滑强度，如图 3 所示：

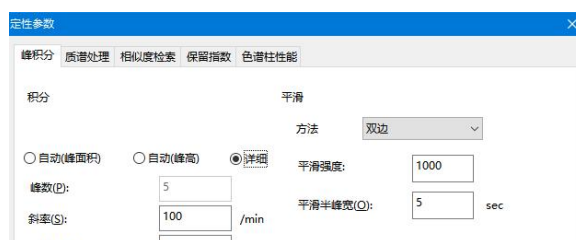


图3 定性积分参数

将平滑方法改为“无”，色谱图恢复正常，如图 4 所示：

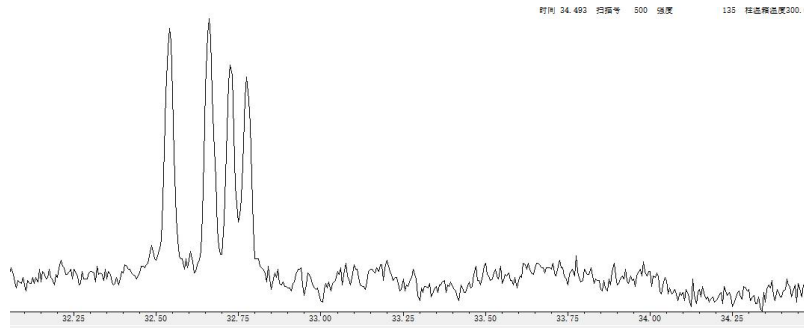


图 4 去除平滑参数的谱图

3 原理解析

数据平滑是提高色谱峰响应信噪比的工作站手段。

我们总是希望得到高质量的色谱信号——良好的色谱分离度，较大的信噪比。然而在分析条件一定的情况下，并不总是能够得到，那么工作站的平滑技术（主要是抑制噪声），就显现出其必要性了。

常见的平滑技术是基于经典数字滤波理论的。根据数学理论，任何一个满足一定条件的信号，都可以被看成是由无限个正弦波叠加而成。或者说，色谱信号可以看作是由不同频率的多个信号叠加而成。

噪声信号大多分布在高频率段，有用的色谱峰信号，大部分分布在较低频率段。那么就可以采取一定的数学算法，将高频段的噪声信号加以滤出或者抑制。从而得到较好的信噪比，即为滤波。

如图 5 所示，图中黑色实线表示的谱图是由红色、绿色、蓝色三条虚线所示谱图叠加而成。红色的信号，可以视为噪声信号。图 6 为该信号的频域图。

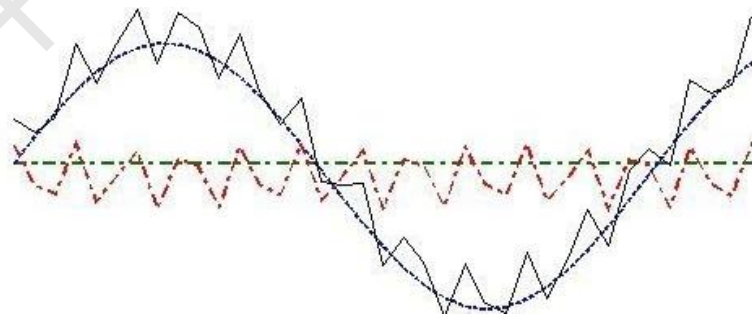


图 5 原始信号时域图

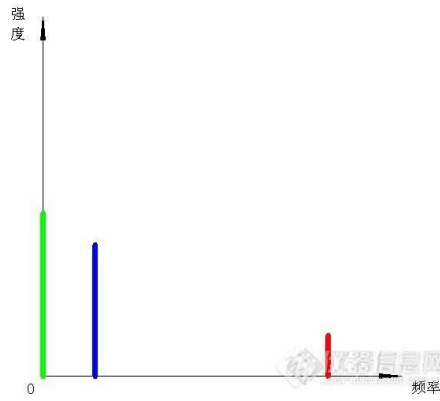


图 6 原始信号频域图

假设我们采用滤波的办法，抑制红色的高频信号，那么就可以得到如图 7 所示的平滑谱图。

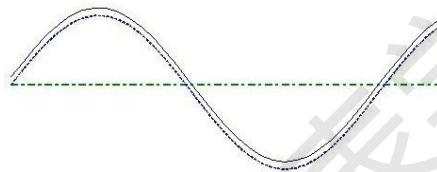


图 7 平滑之后的信号

当然，实际的情况要远比图示中的例子复杂，噪声和信号的频率范围也是比较宽的，但是基本原理是这样的。

GCMSsolution 工作站使用了两种数据平滑的方法，移动平均法和最小二乘法。

3.1 移动平均法（双边法）

移动平均法的原理，确定一定的间隔（窗口），将窗口内的数据点取平均值，然后移动到下一个点。然后重复此过程，直至完成全谱图的处理。

实例：例如设定间隔为 7，那么图中的第 1 至第 7 点进行平均，替换第 4 点。然后窗口向后移动，从第 2 点至第 8 点，重复此计算过程，直至数据结束。如图 7 所示：

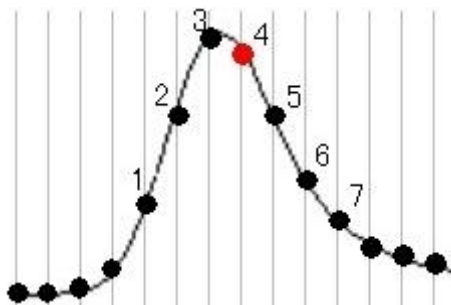


图 8 移动平均法图示

窗口选择越大，滤波效率越好，但是会造成较大的色谱峰畸变。窗口一般选择小于峰宽的 1/10。

有些工作站也提供加权的移动平均平滑法，如图，距离第 4 点越远的数据点，对平滑的贡献越小，计算中赋予较低的权重。同时考虑数据点强度的权重的方法，也称双边法平滑。

LCsolution 的 Smooth 功能，也是使用了移动平均法。滤波效果如图所示：

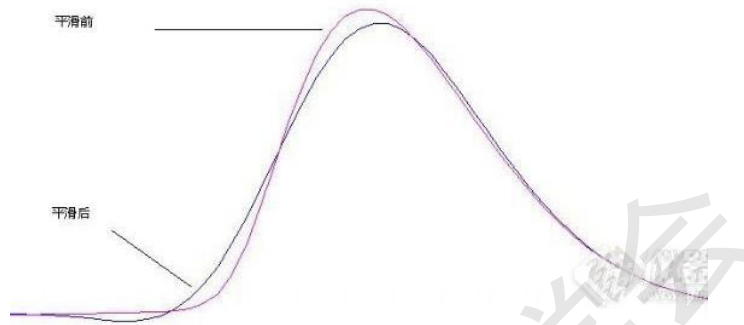


图 9 色谱信号平滑前后的比较

3.2 最小二乘法（Savitzky-Golay 方法）

采用构造多项式来拟合色谱信号的方法，平滑色谱图。

假设一组数据点服从一个多项式，根据最小二乘的原理，定出多项式中各项系数。然后用多项式的值代替实验值，实现平滑。

常用的办法是五点二次平滑和七点三次拟合。

算法如下：

七点三次平滑：

$$y_i = \frac{7x_i + 6 \times (x_{i+1} + x_{i-1}) + 3 \times (x_{i+2} + x_{i-2}) - 2 \times (x_{i+3} + x_{i-3})}{21}$$

五点二次平滑：

$$y_i = \frac{-3x_{i-2} + 12x_{i-1} + 17x_i + 12x_{i+1} - 3x_{i+2}}{35}$$

总体上最小二乘法滤波的效果要好于移动平均法。

4 小结

注意：平滑本质上会丢失原始色谱信号的细节，需要谨慎选择平滑强度。

在一定的范围内增加平滑强度，会改善色谱峰信噪比（这是极为常见的数据处理方法），但如果平滑强度太大，会使得色谱信号高频部分丢失严重。表现为色谱峰峰宽增加，相邻色谱峰的分离度降低，峰高降低，反而会降低信号的信噪比。